Integrating QMR with a Computer-Based Patient Record

AM van Ginneken, EB Liem, P.W Moorman
Dept. of Medical Informatics, Erasmus University
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

ABSTRACT

Most diagnostic decision support (DDS) systems are used as stand-alone applications. At present, the physician can only benefit from the suggestions of a DDS system if he is sufficiently motivated to re-enter patient data and run a diagnostic case analysis. If data from a computer-based patient record (CBPR) could be made available electronically to a DDS system, the use of that DDS system may become much more practical. Integrating a CBPR with a DDS system requires a mapping between two different data structures and dictionaries. We have explored a strategy to create a mapping between our CBPR and QMR. Our research has provided more general insight in the potential and limitations of such a mapping.

INTRODUCTION

Several diagnostic decision-support (DDS) systems have met with good evaluations with respect to the diagnostic performance within their domain, but they also have their limitations [1-4]. Most of them are used as stand-alone applications. As a consequence, the use of a DDS system requires data to be extracted from a patient case and to be re-entered. The user-interface is usually designed to ensure data-entry in a suitable format and may be time-consuming. Furthermore, the use of a stand-alone application depends on the initiative of the physician: he will only consider to use it when he expects to benefit from it. The potential of a DDS to detect possible diagnostic omissions or inconsistencies, independent of the physician's initiative cannot be exploited. One may consider to use a DDS system for both diagnostic support and record keeping to eliminate the extra effort of data-extraction and -entry. However, the domain of the DDS system would limit the expression capability of the physician and its inference processes may even introduce a bias in data collection.

Integration between the CBPR and a DDS system has the advantage that data need not be entered twice and that data-entry can be independent of any DDS system. However, such an integration requires a two-way mapping between the CBPR and the DDS system to be integrated. Firstly, patient data must be represented in a format, suitable for interpretation by the DDS system, and secondly, it is desirable to add to the CBPR all extra data that the DDS system elicits from the physician during case analysis.

Many DDS systems have their own formal representation of patient data. We will be using the term 'data format' for all requirements that patient data have to meet to be suitable for analysis by a DDS system. The better one data format can be expressed in terms of another, the more meaningful a mapping becomes. There has been research involving mapping between clinical vocabularies. In the study by Masarie et al. an interlingua was developed to map between QMR, DXplain, HELP PTXT, and MeSH [5]. The role of our CBPR in the mapping comes closest to that of the interlingua in the aforementioned study. In general, if a CBPR would be complete and unambiguous, then any mapping can be defined for the integration of a DDS system. Besides our efforts towards such a CBPR [6,7], we are currently studying what is involved in mapping cases between our CBPR and QMR [8]. The results helped us understand better to which extent adequate mapping can be achieved without human intervention and when extra information has to be provided to optimize the mapping. In the following, we will explain our strategy and discuss its strengths and limitations.

BASIC CONSIDERATIONS

Our CBPR and QMR have in common that a case consists of a collection of findings, that are explicitly present or absent in the patient. However, patient data are represented in a different format in these two applications. A patient case in the CBPR consists of a tree of atomic medical concepts. With the aid of a flexible data-entry program the physician traverses the knowledge base, thereby creating the structured patient case. The CBPR application supports a fine granularity and the number of meaningful combinations to describe different complaints and findings is beyond count. In QMR, the number of terms, denoting signs and symptoms, is very large (appr. 5000), but fixed. A patient case consists of a list of selected QMR-terms, many of which involve combinations of medical concepts. Such composite findings in QMR would involve small trees when expressed in the CBPR. Hence, a mapping between QMR and our CBPR involves the two different vocabularies as well as the two different data models.

We have chosen to start with the mapping from QMR to our CBPR for two reasons. Firstly, the number of possible combinations to describe findings in the CBPR is very large and not fixed. Part of the findings that can be described in the CBPR have no equivalent in QMR. Although the vocabulary of QMR is large, it contains findings within a diagnostic domain, whereas a CBPR is intended to capture any information pertaining to a patient. A complaint like "nausea" is not present in QMR's finding dictionary, because it is very nonspecific information. In a CBPR, however, a physician must be able to record such a complaint. Since the knowledge base of the CBPR-application can be changed and extended in an easy and flexible way with the aid of a knowledge editor, every QMR-term can be expressed in the data format of the CBPR. For research purposes, we have restricted the mapping to the QMR-findings, involving auscultation of the heart.
The second reason is that the CBPR-expressions of QMR-terms can serve as a basis for mapping the other way: a CBPR-finding can be matched with expressions of QMR-terms in the same data format. Having dealt with the mapping of the data format in the first step (from QMR to the CBPR), the second step focuses on the actual matching. However, even when the data formats are the same, many findings will not result in an exact match, but a partial one. We have explored ways to deal with such matching problems and we explain these in the next section.

MAPPING FROM QMR TO THE CBPR

The mapping from QMR to CBPR involves the definition of CBPR-expressions of QMR-terms. The mapping has to be defined once and only needs revision when QMR-terms are added or changed. Here, we will explain how the initial definition has been realized. For each QMR-term, the user specified the CBPR-expression with the aid of the data-entry program of CBPR-application. When necessary, we added medical concepts to the knowledge base of the CBPR-application. The resulting CBPR-expression consisted of a tree of medical concepts.
Part of the QMR-terms denoted the absence of a finding. We will refer to such QMR-terms as "negative QMR-terms". In case of a negative QMR-term, the CBPR-expression received the marker "absent".
We discovered almost immediately that some QMR-terms cover a set of more detailed findings and, consequently, there is more than one CBPR-expression that justifies a positive match. For example, the QMR-term "murmur continuous in the second or third interspace left" is true when it is heard in the second interspace left, the third interspace left, or both. For such QMR-terms, we defined more than one CBPR-expression.

MAPPING FROM THE CBPR TO QMR

Matching a CBPR-case with QMR-terms involves a match of each finding in that case with each QMR-term (in our restricted set). We wanted an algorithm that would allow us to mark each QMR-term, that is matched, as "True", "False", or no mark, meaning "Unknown". Those QMR-terms that are marked True or False make up the QMR-expression of the CBPR-case. In the following, we focus on the strategy for the match of one CBPR-finding with a CBPR-expression of a QMR-term. The match of a whole CBPR-case with QMR involves a repetition of this strategy.

Matching criteria

Prior to designing a matching strategy, we had to define criteria for marking QMR-terms as True or False. It is evident that a QMR-term is True when its CBPR-expression is identical to a CBPR-finding. However, we cannot state that a QMR-term is False otherwise. It depends on the QMR-term being positive or negative in combination with the absence or presence of the CBPR-finding how the results of partial matches are to be interpreted. When the QMR-term is positive and the CBPR-finding is present, then that QMR is True when at least one if its CBPR-expressions is True. A CBPR-expression of that QMR-term is True when it is equally or less detailed than the CBPR-finding. For example: when a murmur is present in the second interspace, then the QMR-term "murmur present" is True. When the QMR-term is negative, different rules apply to decide about the result of the match. As a consequence, the matching strategy must be able to discriminate among the following possibilities: the CBPR-finding is identical to, less detailed than, or more detailed than the CBPR-expression of a QMR-term. In addition, the matching strategy must include criteria how to interpret these possibilities in combination with positive or negative QMR-terms and present or absent CBPR-findings.

An intermediate data format: Pathways

In order to establish whether a CBPR-finding is identical to, less detailed than, or more detailed than a CBPR-expression of a QMR-term, we decided to use an intermediate representation: each CBPR-tree is broken down into separate pathways.

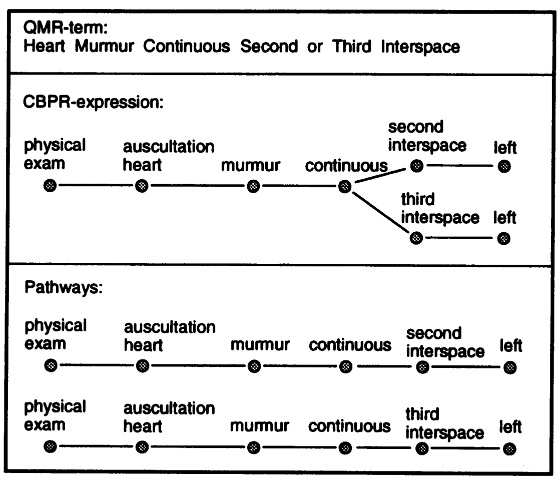

Figure 1: The CBPR-expression of a QMR-term and its decomposition in pathways.

A pathway consists of a sequence of medical concepts from the root to one of the leaves.

Each CBPR-finding can now be represented by a set of one or more pathways, which may be marked as absent. The same holds for a CBPR-expression of a QMR-term. Since there may be several CBPR-expressions for one QMR-term, a QMR-term may be represented by one or more sets of pathways. An example of a QMR-term and its set of pathways is shown in Figure 1.

Matching

All pathways in all sets, defined for a QMR-term, are compared with the pathways of a CBPR-finding. We will refer to these pathways as "QMR-pathways" and "finding-pathways" respectively. The first step of the match determines for each possible combination whether the finding-pathway is identical to, longer (more detailed) than, or shorter (less detailed) than the QMR-pathway.

To determine for each QMR-pathway if it is True or False, the second step of the match takes into account the positive or negative definition of the QMR-term and the presence or absence of the CBPR-finding. For example, when the QMR-term is positive and the CBPR-finding is present, then a QMR-pathway is True when the finding-pathway contains the QMR-pathway, i.e. when the finding-pathway is identical or longer then the QMR-pathway. In case of an absent finding-pathway, a positive QMR-pathway is False when the QMR-pathway contains the finding pathway. In other words: when a murmur is absent in a patient, then all QMR-pathways including a murmur are False.

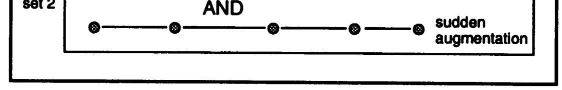

Figure 2: The representation of a QMR-term in sets and pathways with the Boolean-operators that have to be applied when the QMR-term is positive and the finding is present. The QMR-term is TRUE when one of its sets is TRUE (OR) and a set is TRUE when all its pathways are TRUE (AND).

Having established which QMR-pathways are True and which are False, the matching algorithm must determine whether or not the QMR-term as a whole is True or False. Here too, the positive or negative definition of the QMR-term and the presence or absence of the CBPR-finding are important. A positive QMR-term may be marked as True when at least one of its CBPR-expressions is True, and such an expression is True when all its pathways in the corresponding set are True. In other words: the AND operator applies to pathways within a set and the (inclusive) OR operator applies to the sets of the QMR-term (see Figure 2). Different rules apply when the QMR-term is negative, or when the CBPR-finding is absent.

	QMR-finding positive	QMR-finding negative
CPBR-finding present	Q-path TRUE if: F-path $\supseteq$ Q-path QMR-set TRUE if: ALL paths are TRUE (AND) QMR-term TRUE if: $\geq$ 1 set is TRUE (OR)	Q-path FALSE if: F-path $\supseteq$ Q-path QMR-set FALSE if: $\geq$ 1 path is FALSE (OR) QMR-term FALSE if: ALL sets are FALSE (AND)
CBPR-finding absent	Q-path FALSE if: F-path $\subseteq$ Q-path QMR-set FALSE if: $\geq$ 1 path is FALSE (OR) QMR-term FALSE if: ALL sets are FALSE (AND)	Q-path TRUE if: F-path $\subseteq$ Q-path QMR-set TRUE if: ALL paths are TRUE (AND) QMR-term TRUE if: $\geq$ 1 set is TRUE (OR)

Figure 3: The inclusion criteria for pathways and the Boolean-operators that have to be applied to determine whether a QMR-term is TRUE or FALSE. The upper left quadrant applies to the example of Figure 2.
(Q=QMR, F=Finding, $\subseteq$ = identical or shorter)

Figure 3 shows for all 4 possible combinations of positive and negative with the presence and absence of a finding, how pathways and sets are matched to assign TRUE or FALSE to the QMR-term. The match results in Unknown when none of the situations applies.

When applying the above mentioned strategy sequentially to all findings of a case, a QMR-term may receive many FALSE and a few TRUE assignments. So far, we have used the following rule for matching a whole case: 1) a QMR-term is TRUE when at least one match with a CBPR-finding returns TRUE and 2) a QMR-term is FALSE when all matches return FALSE. The problem of conflicting evidence in a patient case is subject for further research.

DISCUSSION

We have been able to realize the essence of a mapping from QMR to the CBPR and vice-versa. This research contributed much to our insight in the potential of integrating expert systems with CBPR's. The most important advantages of integration that motivated our research are the reduction of effort for data-entry and, even more, the possibility to benefit from DDS systems, independent of the physician's initiative.
Masarie et al (5) found that an interlingua needs to be exhaustive with respect to the contents of all vocabularies that are to be included in a mapping. This can be achieved by defining new generic frames. Since the CBPR is the source of clinical data, an interlingua will be most effective when it can represent the findings in such a CBPR or when that CBPR can serve as interlingua itself. Generic frames as described in the study do not seem to have the same expression capability as our CBPR. A paper describing the formal structure of our CBPR in detail is approaching its completion.
The most important finding in our study was that, within the domain of auscultation of the heart, many QMR-terms remained unknown, due to unresolved partial matches. The problem with automated interpretation of patient data is that the physician is not aware of which findings will be taken into account by the DDS system. It is likely that the physician expects findings, within the domain of the DDS system, to be subjected to the inference process, but this may not reality. This problem would not occur with stand-alone use where the physician directly decides about True or False. Unresolved partial matches result from the difference between judging predefined statements and the freedom to actively specify findings. When actively entering findings in a CBPR, the physician may express a finding with less or more detail than one formulated in a DDS system. Most likely, the physician has the knowledge needed to resolve the match. However, the practical advantage of automated interpretation is lost when it is accompanied by many questions to the physician. The fundamental question is how good a mapping can be ultimately achieved without human intervention: to what extent can we infer part of the missing information and if there is information that will always require human intervention, what is it's nature?

Since it was not difficult to map findings from QMR to the CBPR, it is, in principle, easy to add new findings to the CBPR, which are the result of a case-analysis session with QMR. However, the physician may deliberately influence the result of the system's reasoning by marking findings as positive, which have not been confirmed in the patient. For example, a DDS system may have a finding like "murmur systolic in the third interspace", but not a finding like "murmur systolic in the fourth interspace". When a physician hears a systolic murmur in the fourth interspace, he may choose to mark the finding in the third interspace as present. In other words, the physician may mark certain findings when he expects them to make the case analysis more realistic, even when these findings have not been established in the patient. It is not desirable to add to the CBPR findings that have not been actually found in the patient. On the other hand, verification of all extra findings after a case analysis, reintroduces repetition of data-entry. We conclude that the problem of unresolved partial matches as well as the entry of findings for the purpose of realistic case analysis, all concern findings that come close to ones that have actually been found in the patient. We considered it desirable that the matching strategy can detect when these problems may occur, i.e. identify patient findings that have much in common, but are different from those in the DDS system. Our intermediate format and matching strategy allows for this detection. Additional research is required to find the most efficient way to deal with these situations.

Although we have only taken the first steps on the way to integration of a DDS with a CBPR, we are beginning to understand some fundamental problems involved with such an integration. Our methodology, especially the representation in pathways and the matching strategy may have aspects that are applicable to integration of CBPR's and DDS's in general. Because of the limited domain of a DDS, because no CBPR can replace the physician's knowledge about his patient, and because the physician is a very powerful reasoner, he should always make his differential diagnosis prior to consulting a DDS system. In the near future, we will focus on the value of the integration of CBPR's with DDS systems for offering suggestions as a supplement to the physician's own differential diagnosis.

Reference

[1] Duda RO, Shortliffe EH. Expert Systems Research. Science 1983;220:261-268.

[2] Miller RA. INTERNIST-1/CADUCEUS: Problems Facing Expert Consultant Programs. Meth Inform Med, 1984;23:9-14.

[3] First MB, Soffer LJ, Miller RA. Quick (Quick Index to Caduceus Knowledge): Using the Internist-1/Caduceus Knowledge Base as an Electronic Textbook of Medicine. Comp Biomed Res, 1985;18:137-165.

[4] Bergeron B. Iliad: a diagnostic consultant and patient simulator. M.D. Computing, 191;8:46.

[5] Masarie FE, Miller RA, Bouhaddou O, et al. An Interlingua for Electronic Interchange of Medical Information: Using Frames to Map between Clinical Vocabularies. Comp Biomed Res, 1991;24:379-400.

[6] van Ginneken AM, van der Lei J, Moorman PW. Towards Unambiguous Representation of Patient Data. In: Frisse ME ed. Proceedings of 16th Annual Symposium on Computer Applications in Medical Care, Baltimore 1992:69-73.

[7] Boon WM, Duisterhout JS, van Ginneken AM. New functional requirements for Electronic Medical Records. In: Lun KC et al. eds. Proc. MEDINFO 92. Elsevier Science Publishers B.V. (North Holland), 1992:697-702.

[8] Parker RC, Miller RA. Creation of Realistic Appearing Simulated Patient Cases Using the INTERNIST-1/QMR Knowledge Base and Interrelationship Properties of Manifestations. Meth Inform Med, 1989;28:346-351.